DOCUMENT RESUME

ED 128 011 IR 003 982

AUTHOR Donati, Robert
TITLE Survey of Online Access to Social Science Data Bases.
PUB DATE 8 Jun 76
NOTE 21p.; Paper presented at the Special Libraries Association Annual Conference (Denver, Colorado, June 8, 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Data Bases; *Information Retrieval; *Information Sources; *On Line Systems; *Social Sciences; Surveys
IDENTIFIERS *Lockheed DIALOG

ABSTRACT

Until very recently there was little computer access to comprehensive bibliographic data bases in the social sciences. Now online searching of several directly relevant files is made possible through services such as the Lockheed DIALOG system. These data bases are briefly surveyed, with emphasis on content, structure, and strategy appropriate for online interactive searching. Indexes discussed in this paper include Social Science Citation Index, Sociological Abstracts, Psychological Abstracts, Language and Language Behavior Abstracts, Historical Abstracts, Educational Resources Information Center (ERIC), Exceptional Child Education Abstracts, Foundations Directory, Foundations Grant Index, and others. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

SCOPE OF INTEREST NOTICE
The ERIC Facility has assigned this document for processing to: IR 50
In our judgement, this document is also of interest to the clearinghouses noted to the right. Indexing should reflect their special points of view.

ED128011

SURVEY OF ONLINE ACCESS TO SOCIAL SCIENCE DATA PASES

By Robert Donati
Lockheed Information Systems
New York, N. Y.

Abstract

Until very recently there was little computer access to comprehensive bibliographic data bases in the social sciences. Now online searching of several directly relevant files is made possible through services such as the Lockheed DIALOG(R) system. These data bases are briefly surveyed, with emphasis on content, structure, and strategy appropriate for online, interactive searching. Indexes discussed in the paper include Social Sciences Citation Index(R), Sociological Abstracts, Psychological Abstracts, Language and Language Behavior Abstracts, Historical Abstracts, ERIC, Exceptional Child Education Abstracts, Foundations Directory, Foundations Grants Index and others. Coverage of certain social science topics is quantitatively compared among several social science and more general files, including NTIS and Comprehensive Dissertations Abstracts Index. Techniques for online thesaurus utilization are described as are systematic application of the same strategy across files through a search save feature and the use of merged keyword and term indexes from several data bases. The relatively modest costs of such services are briefly analyzed.

Given June 8, 1976 on the panel "Data Base Update: Innovations in Social Science Information Handling," during the Social Science Division session of the 67th Annual Convention of the Special Libraries Association, Denver, Colorado.

IR003982

Social science researchers and librarians need no longer feel "left out" of the rapid developments which have occurred in computer-based information systems. Within the last two years, many tools of information science have been applied to the social sciences so that a substantial amount of information is now more widely accessible.

For example, there are now eight social science bibliographic data bases on the Lockheed DIALOG(R) online retrieval service totalling approximately 1,070,000 document records as of early June, 1976. Furthermore, perhaps 25% of over 1.1 million additional records in six closely related files provide further in-depth information of some interest to the social scientist. Thus, over 1.3 million document records directly relevant to the social sciences are available - representing 10% of the 13 million records loaded on DIALOG. This percentage cannot be regarded as small, given the extensive size of the scientific literature and the long-established, government-funded programs in scientific and technical information.

Just what data bases are now available? With respect to the social sciences, there are both multidisciplinary files such as Social Sciences Citation Index(R) and Comprehensive Dissertation Abstracts Index, and a number of discipline oriented files such as Psychological Abstracts, Sociological Abstracts, and Historical Abstracts. A complete list of those now available on the DIALOG service is given in Figure 1. There are also other data bases, such as ABI/INFORM and NTIS, which have high social science content. Many of the scientific

ED 003 848

Figure 1

DIALOG(R) DATA BASES IN SOCIAL SCIENCES AND RELATED AREAS

Data Base Name	Subject and Source	Content and Currency: Coverage (Entry Year To Present)	No. of Citations (4/1/76)	Update Freq'y	Citations Added
Social Sciences					
SOCIAL SCISEARCH(R)	Multidisciplinary index covering the world's most important social science journals; from the Institute for Scientific Information	Jan 1972	367,500	Monthly	7,500
PSYCHOLOGICAL ABSTRACTS	World literature in psychology and other behavioral sciences; from the American Psychological Association	Jan 1967	230,000	Monthly	2,000
SOCIOLOGICAL ABSTRACTS	In-depth coverage of sociology and related social science areas; from Sociological Abstracts, Inc.	Jan 1963	70,000	Quarterly	2,000
LANGUAGE AND LANGUAGE BEHAVIOR ABSTRACTS	World literature on speech and language pathology; from Sociological Abstracts, Inc.	Jan 1965	30,000	Quarterly	1,000

Figure 1, (cont'd)

ERIC	Research and journals in the field of education; from the Educational Resources Information Center	Jan 1966	242,500	Monthly	2,500
EXCEPTIONAL CHILD EDUCATION ABSTRACTS	Education of handicapped and gifted children; from the Council for Exceptional Children	Jan 1966	21,250	Quarterly	750
HISTORICAL ABSTRACTS	Worldwide history from 1450 to date. See below for U.S. and Canada; from American Bibliographic Center, Clio Press, Inc.	1973-75	23,600	Monthly	625
AMERICA: HISTORY & LIFE	U.S. and Canadian history, social life, patterns, urban affairs, etc. Includes journal articles; from American Bibliographic Center, Clio Press, Inc.	1955-74	55,900	Monthly	900

Figure 1, (cont'd)

Other Relevant Files

COMPREHENSIVE DISSERTATION ABSTRACTS INDEX	Doctoral dissertations from accredited universities (predominantly U.S.); from Xerox University Microfilms	1861	530,000	Monthly	3,300
NTIS	Complete Government Reports Announcements data base covering a broad range of disciplines of over 240 government agencies; from the National Technical Information Service	Jan 1964	509,500	Biweekly	2,300
FOUNDATION DIRECTORY	Descriptions of over 2,500 foundations with assets exceeding $1 million; from the Foundation Center	Current	2,500	Semi-annually	Compl. Rplct.
FOUNDATION GRANTS INDEX	Cumulation of grants records of more than 400 U.S. philanthropic foundations; from the Foundation Center	Jan 1973	27,000	Bi-monthly	200
ABI/INFORM	Abstracted Business Information covering business, finance, and related fields; from ABI, Inc.	Aug 1971	34,200	Monthly	800

and business files also contain material of interest. For example, the field of linguistics is especially well covered in INSPEC Computers and Control Abstracts.

The question of database selection for a particular search is, I believe, a less difficult choice than it often appears. There are now available at least three and frequently several more good choices for virtually any topic in the social sciences: (1) the appropriate functional database, such as Language and Language/ Behavior Abstracts, (2) the only truly comprehensive social science database - Social Sciences Citation Index(R), and (3) Comprehensive Dissertations Index. ERIC is almost always a good fourth choice, especially where the teaching of the subject is involved.

As an aid to data base selection, I have prepared a rudimentary directory of subject content for the files directly concerned (Figure 2). The subject areas include the 25 social science topics chosen for the informal roundtable discussions at the meetings of the SLA Social Science Division and were augmented by a few additional topics. For each data base I rated coverage of each topic as high (H), moderate (M), small (S), or minor/negligible. These opinions are my own based on a review of the subject classification schemes, journal coverage, and my own experience with the files. Unfortunately, a lengthy study would be required to make a more rigorous quantification of content. I would welcome comments from librarians and abstracting and indexing services on the validity

Figure 2

SUBJECT DIRECTORY To DATA BASES	Soc.Scisearch (R)	Psych. Abs.	Sociol.Abs.	Lang/Lang Behavior	ERIC	Exc. Child Abstracts	Histor.Abs.	America Life & Hist.	Comp. Dissert. Index	NTIS	Foundation Dir.& Grants	ABI/INFORM	OTHER DATA BASES
Anthropology/ Archaeology	M						S	S	S		S		
Business & Finance	S				S				H	M	S	H	Predicasts
Communication	S	M	M	M	M	S			H	H	M	H	Predicasts, INSPEC, EI
Crime & Law Enforcement	H	M	S		S	S	S	S	S	H	M	M	
Drug Abuse & Alcoholism	M	H			M	M				M	S	M	BIOSIS
Economics	H				M		M	M	H	M	S	H	All Predicasts files
Education	H	H			H	H	S	S	H	M	H	S	
Elderly & Handicapped	H	H	S		S	S			S	M	M	S	BIOSIS
Employment & Labor	S	M	S				S	S	M	M		H	Predicasts
Foundations							S	S	M	S	H	S	
Governments	M		S				[illegible]	H	H	M		M	
Health Care	H	M			M	M			H	H	H	S	CAIN, BIOSIS, PREDICST
History	H						H	H	H		S		Eng.Irdex, CAIN
Housing & Urban Dev.	M	M			S		S	S	M	H	S	S	

SUBJECT DIRECTORY To DATA BASES	Soc.Science (R)	Psych. Abs.	Sociol. Abs.	Lang/Lang Behavior	ERIC	Exc. Child Abstracts	Histor. Abs.	America Life & Hist.	Comp. Dissert. Index	NTIS	Foundation Dir. & Grants	ABI/INFORM	OTHER DATA BASES
International & Foreign	M						H	H	H	S	H	M	
Legislative & Legal		H	M		S	S	S		H	M	S	H	CAIN
Linguistics		M	M	H	M	S			S				INSPEC
Personnel & Management		M	M	M	S				H	M	S	H	ABI/INFO M, EI
Philosophy		M					S	S	H		S		
Planning & Zoning							S	S	M	H		S	Engr.Index,CAIN
Psychology, Mental Health		H	H	S	M	M			H	M	M	S	
Religion		S	M				M	M	H		S		
Social/Econom. Statistics		S			S		S	S	M	S		H	Pred.Stat.,CAIN
Social Issues		H	M	H	H	S	H	H	H	M	H	M	CAIN
Social Welfare & Work		M	M	S	M	S	S	S	M	H	M	S	CAIN
Sociology		H	M	H	M	S			M	M	M		
Urban Affairs		M	M	S	M	S	M	M	M	H	M	S	Engr. Index
Women's Issues		M	M	S	H		S	S	M		S	S	CAIN

of the chart. I believe, however, the table to be especially useful in steering the novice user to the more important data bases which cover a given topic.

Which files need to be used depends, of course, on the purpose and scope of the search. Primary developments in the field might be covered by examining one or two discipline-oriented data bases. If government activity or sponsorship is involved, then NTIS should probably be considered as well. A question involving behavior of individuals, groups, organizations, should likely be searched in Psychological Abstracts. Information on current research may frequently be found in Foundation Grants Index. If an extremely comprehensive bibliography is required, it may be necessary to search a half-dozen or more files.

To obtain a more precise evaluation of coverage, a search was run on several key data bases to determine the number of document records in which each phrase occurred literally. The results (Figure 3) again show multiple data base coverage for many topics.

Most of the data bases with which we are concerned are bibliographic, wherein the machine record is a surrogate for the original document containing at least title, authors and bibliographic citation. A review of the major data elements of the DIALOG social science data bases reveals a great similarity (Figure 4) in structure with but few significant exceptions. Most of the files also contain assigned subject indexes (controlled and uncontrolled), narrative

Figure 3

OCCURRENCE OF SELECTED PHRASES IN CERTAIN

DIALOG SOCIAL SCIENCE DATA BASES

Exact Phrase	ERIC	NTIS	SOCIAL SCISEARCH (R)	PSYCH ABSTRACTS	ABI INFORM	FOUNDATION GRANTS IDX	DISSERT. INDEX
leisure time	561	16	24	153	54	0	58
leisure activities	53	2	11	43	8	0	13
social history	82	1	201	43	2	2	112
social welfare	395	1,019	154	158	43	960	79
social indicators	64	56	72	17	14	1	13
municipal finance	1	3	4	0	4	0	13
urban affairs	94	26	97	6	18	41	4
urban studies	321	109	136	4	2	56	4
economic controls	2	5	1	0	13	0	1
gasoline shortage	0	0	1	0	19	0	0
national health insurance	16	15	39	27	88	1	5
no fault insurance	1	1	19	0	34	1	0
world food	132	13	44	6	78	16	8
music therapy	22	0	13	125	0	5	5

Figure 4

MAJOR DATA ELEMENTS OF DIALOG SOCIAL SCIENCE DATA BASES

Data Base	Title	Auth.	Auth. Affil.	Journ. Cits.	Cited Auth.	Assigned Indexes Contr.	Assigned Indexes Uncontr.	Thesaurus On Line	Numer. Classif. Scheme	Abst.
SOCIAL SCISEARCH	x	x	x	x	x					
PSYCH ABS	x	x	x	x		x	x	x	x	x
SOCIOLOG ABS	x	x	x	x		x	x		x	x
LANG/LB ABS	x	x	x	x		x	x		x	x
ERIC	x	x	x	x		x	x	x		x
EXC CHILD ABS	x	x	x	x		x	x	x		x
HISTOR ABS	x	x		x		x	x		x	x
AMER HIST & LIFE	x	x		x		x	x			x
DISSERT ABS	x	x	x			x				
NTIS	x	x	x			x	x		x	x
ABI/INFORM	x	x		x						x

abstracts or annotations, and author affiliations. Some also have thesauri online and numerical classification schemes.

There are also several files which are statistical or mixed bibliographic/statistical. Examples of the first are Foundations Directory and Foundations Grants Index; examples of the latter include the Predicasts Market Abstracts, F & S Indexes, and Domestic/International Statistics. Online access is rapidly increasing in data bases which are factual or whose records may be described as "informative" in contrast to the "descriptive" records of the bibliographic data bases.

What effect does knowledge of record structure and content have on search comprehensiveness and precision? The obvious answer may be "plenty." While a good understanding of the files is desirable and usually necessary for high recall of documents and efficient use of time and resources, the advent of online, interactive searching has made it possible to do relative effective searching without this detailed knowledge. Why? -- for several fundamental reasons:

(1) the general ability to interact with results, changing search strategy as one proceeds,

(2) natural language (or "full-text") searching allowing one to look for exact phrases within subject-indicating fields such as titles, indexes, abstracts, etc.,

(3) proximity and field specifications of terms, giving more precision than the simple logical intersection ("and" operation),

(4) display of inverted indexes showing all terms actually used, thus guiding the user to alphabetically and conceptually near terms,

(5) online thesauri which provide expanded cross references to topics,

(6) truncation of word stems, obviating such problems as teenager, teenagers, teen-ager, teen ager, etc..

(7) offline merged term indexes for several data bases by broad subject categories,

(8) Search save capability wherein a search concept (e.g. women's liberation) or an entire search can be defined in full text fashion, saved in the computer, and applied to one or more data bases.

Most of these topics are illustrated in the examples given in Figures 5 and 6. While a searcher should usually divide the topic into logically independent concepts prior to search time, it is really unnecessary always to make an exhaustive analysis of offline search aids if time is short. It is often a wise strategy to take a minimum amount of key input, make logical combinations and then output a few good records online, selecting the significant terms for reincorporation into the strategy.

The search shown in Figure 6 illustrates this point. The topic was "citizen participation in the operation of community facilities." The basic inverted index of the file selected (Psychological Abstracts)

Figure 5

DIALIST(TM)

Term Frequency Indexes from the DIALOG Files

In Microfiche

EXTRACT FROM MERGED-TERM INDEX

SOCIAL SCIENCES GROUP

	ERIC	PSYCH ABSTRACTS	SOCIAL SCISEARCH	EXCEPT CHILD
crime	414	1175	619	31
crime conviction	-	20	-	-
crimes	37	172	101	9
criminal	273	731	997	38
criminal law	7	48	-	-
criminality	13	108	55	6
criminals	109	1081	26	29

Figure 6

Sample DIALOG Search

EXPAND COMMUNITY FACILITIES

Ref	Index-Term	Type	Items	RT
E1	COMMUNITIZED----------		1	
E2	COMMUNITY----------		6075	
E3	COMMUNITY ATTITUDES---------		94	1
E4	COMMUNITY COLLEGE STUDENTS--		10	2
E5	COMMUNITY COLLEGES----------			1
E6	-COMMUNITY FACILITIES--------		51	15
E7	COMMUNITY MENTAL HEALTH-----		76	7
E8	COMMUNITY MENTAL HEALTH			

EXPAND E6

Ref	Index-Term	Type	Items	RT
R1	COMMUNITY FACILITIES-------		51	15
R2	CHILD GUIDANCE CLINICS-----	R	57	7
R3	COMMUNITY MENTAL HEALTH CENTERS----------	N	304	12
R4	COMMUNITY SERVICES	R	1504	8
R5	DAY CARE CENTERS	R	26	3
R6	HALFWAY HOUSES----------	R	29	6
R7	HOUSING----------	N	366	2
R8	PUBLIC TRANSPORTATION------	N	12	4
R9	RECREATION AREAS----------	R	10	4
R10	REHABILITATION CENTERS-----	R	30	3
R11	RELIGIOUS BUILDINGS--------	R		4
R12	SCHOOLS----------	R	4049	15
R13	SHELTERED WORKSHOPS--------	R	55	3
R14	SHOPPING CENTERS----------	N	4	3
R15	SUICIDE PREVENTION CENTERS-	N	30	6
R16	URBAN PLANNING----------	R	20	5

SELECT R1-R6, R9, R10, R13, R15

1 2063 R1-R6, R9, R10, R13, R15
R1: COMMUNITY FACILITIES

SELECT CITIZEN(W)PARTICIPATION

2 20 CITIZEN(W)PARTICIPATION

Figure 6 (cont'd)

COMBINE 1 AND 2

3 11 1 AND 2

TYPE 3/6/1-5

1. DOC YEAR: 1976 VOL NO: 55 ABSTRACT NO: 20739
   A Study of Citizen Participation in a Community Mental Health Center.
2. DOC YEAR: 1976 VOL NO: 55 ABSTRACT NO: 02118
   Citizen Participation in Decisionmaking: Myth or Strategy?
3. DOC YEAR: 1975 VOL NO: 53 ABSTRACT NO: 08054
   Citizen Participation and Conflict.
4. DOC YEAR: 1975 VOL NO: 53 ABSTRACT NO: 02956
   Advocates For Themselves: Citizen Participation in Federally Supported Community Organizations.

TYPE 3/5/4

DOC YEAR: 1975 VOL NO: 53 ABSTRACT NO: 02956
Advocates For Themselves: Citizen Participation in Federally Supported Community Organizations
Moguiof, Melvin B.
Urban Inst, Washington DC
Community Mental Health Journal 1974 Spr Vol 10(1) 66-76
Discusses variations in the intensity of citizen participation in community organizations and variations in the decision structures for participation (e.g., advisory mechanisms or citizen control). It is concluded that although control mechanisms may have certain negative consequences for racial integration, citizen participation should be viewed as a policy goal as well as an instrument for achieving other goals.

CLASSIFICATION- 09
SUBJECT TERMS- COMMUNITY SERVICES, PARTICIPATION: 10690, 36810
INDEX PHRASE-Citizen Participation, Community Organizations

SELECT PARTICIPATION

4 2102 PARTICIPATION

SELECT CITIZEN(W)CONTROL

5 2 CITIZEN(W)CONTROL

COMBINE 1 AND (4 OR 5) NOT 3

6 81 1 AND (4 OR 5) NOT 3

PRINT 3/5; PRINT 6/5/1-81

END

was browsed for "community facilities." This term was used and the computer indicated there were 51 documents posted to it, and 15 cross references. These were then called up and ten appropriate terms were selected, yielding a "set" of 2,063 documents. The exact phrase "citizen participation" was then selected (from the entire data base) and combined with the first group on community facilities. A few titles were typed online. One of the "hits" was reviewed and the subject term "partipation" was noticed as well as the phrase "citizen control." These were then selected, combined with "community facilities" and the additional hits identified. All the hits were then printed in an offline bibliography. This search cost $7.56 for online search time and offline printing for full records of the 92 hits would have been $9.20.

Another example of surprisingly modest costs for online searching is given in Figure 7. Here, a "feminism/women's rights" concept was defined, using eight exact phrases and the truncated stem "feminis---". The strategy was initially defined in Social Sciences Citation Index(R) and then stored away. Five and one quarter minutes were required for this operation. The search was then recalled and executed in each of five other files. A total of 1,466 hits were obtained in all six files (the degree of duplication is unknown but is probably around 15%), requiring a total of 18 minutes terminal time,costing $17.17 in computer and data communication costs.

In assessing the cost/benefits of online searching, one ought

(Figure 7)

SEARCH SAVE EXAMPLE

I. Search Definition and Saving

A general "feminism/womens rights" concepts was defined in the Social Scisearch(R) data base using natural language by selecting and combining (in the "or" sense) all records containing one or more of the following phrases:

feminis--- (truncated form for feminism, feminist, feminists, etc.)

women's rights	women's liberation
womens rights	womens liberation
women's studies	women's lib
womens studies	womens lib

II. Execution of Search on Several Data Bases

	Data Base	No. of Hits	Time (Mins.)	Search Time Cost*
7	Social Scisearch	205	5.22	$ 6.79
11	Psychological Abstracts	159	2.77	2.68
1	ERIC	928	6.15	3.38
15	ABI/INFORM	32	1.13	1.37
27	Foundation Grants Index	35	1.05	1.19
35	Compr. Dissert. Index	107	1.68	1.76
	Totals	1,466**	18.00	$17.17

* Covers data base rate and data communications network. Excludes offline printing.

** This is the total with duplicate citations.

to consider the costs per typical searches rather than search rates. Most searches cost the user somewhere between $2.00 and $20.00. These figures include all costs paid to the retrieval service vendor (computer time, data communications, and offline printing) but exclude terminal rental, telephone cost (if any) to reach a network number, and library personnel time. The average search time determined from hundreds of thousands of searches is currently running 10 minutes although many searches are done in a few minutes and some may require 20 or more minutes. The average number of offline prints is 24. Using these statistical averages, the average search cost in certain data bases is as follows:

ERIC	$7.90	PSYCH ABSTRACTS	$12.07
NTIS	$9.57	SOCIAL SCISEARCH	$15.40

The savings of many man hours to perform such searches manually (or even the impossibility of conducting such searches manually) appear to be substantial when compared to the relatively modest costs of searching indicated above.

In conclusion, it is interesting to observe that in this year which marks the centennial of the telephone and roughly the tenth anniversary of online interactive retrieval, that social scientists and librarians now have at their disposal the mutually harnessed

technology of both communications and computers including

- a large number of discipline oriented files
- several good multi-disciplinary data bases
- online and offline vocabulary identification aids and
- cross database search capabilities.

It is hoped that usage will grow so that future developments in database coordination and retrieval system capabilities can be based on extensive experience of social science librarians and their patrons.